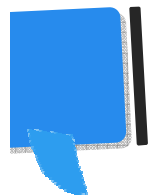


---

# Language Technology: The Missing Link in the Digital Agenda





# LT-Innovate

The Forum for Europe's Language Technology (LT) Industry



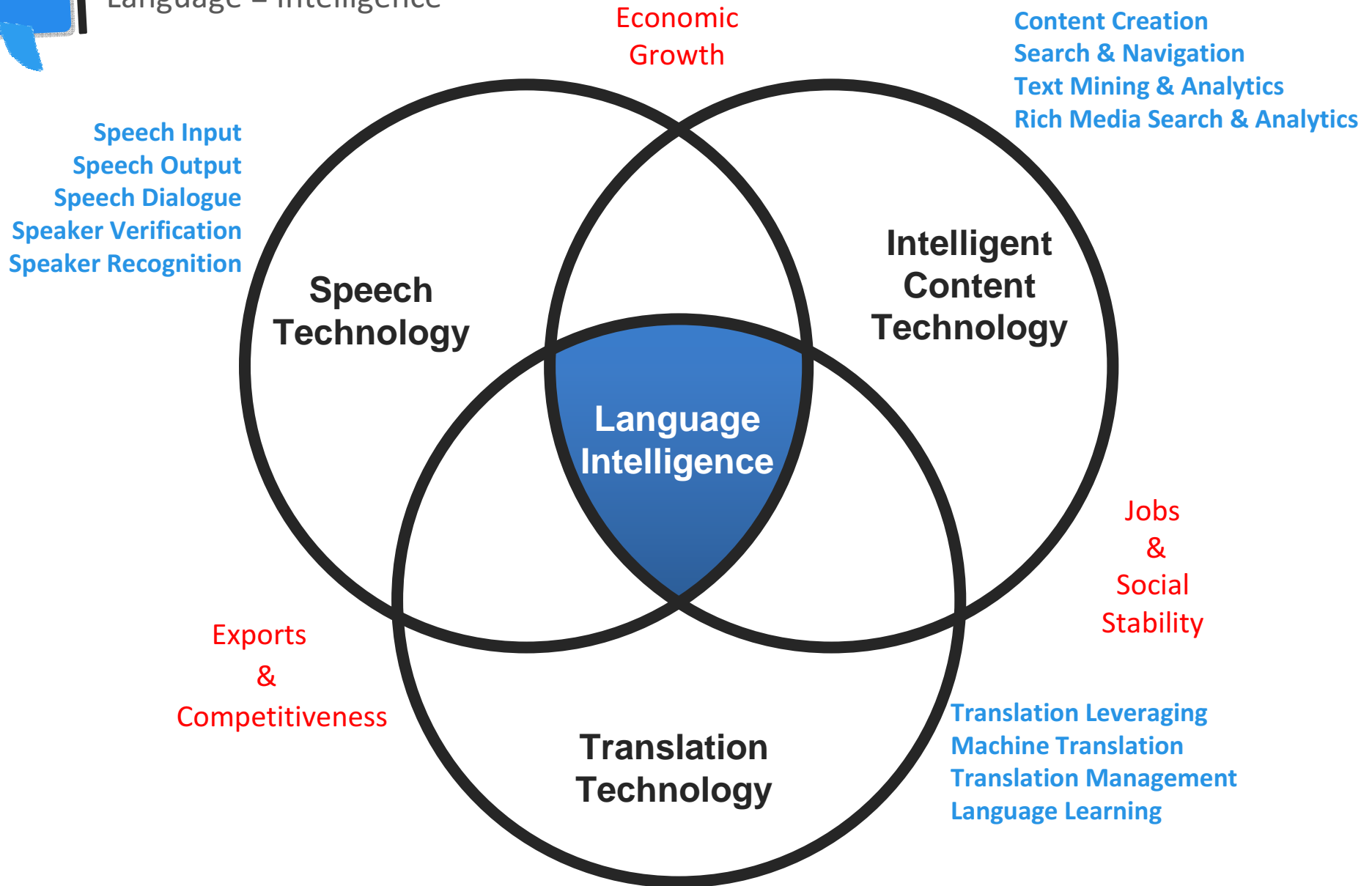
- Created in January 2012 – 90 full members in June 2012
- Gathers the LT Industry (SMEs)
- Develops a common vision (LT Innovation Agenda)
- Assembles and disseminates key LT market data
- Increases the visibility of the LT industry vis-à-vis buyers and policy makers
- Organises the collaboration of LT SMEs and other LT value-chain stakeholders (through Special Interest Groups)
- Facilitates collaborative projects
- Facilitates the transfer of research results and accelerates the uptake of innovation in the market place
- Organises the first LT-Innovate Summit on 19 June in Brussels: [www.lt-innovate.eu/summit](http://www.lt-innovate.eu/summit)
- Chair: [Jochen Hummel](#)
- Secretary General: [Philippe Wacker](#)
- [www.lt-innovate.eu](http://www.lt-innovate.eu)
- [contact@lt-innovate.eu](mailto:contact@lt-innovate.eu)

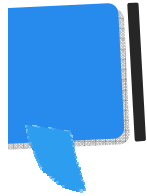




# What Language Technology Does

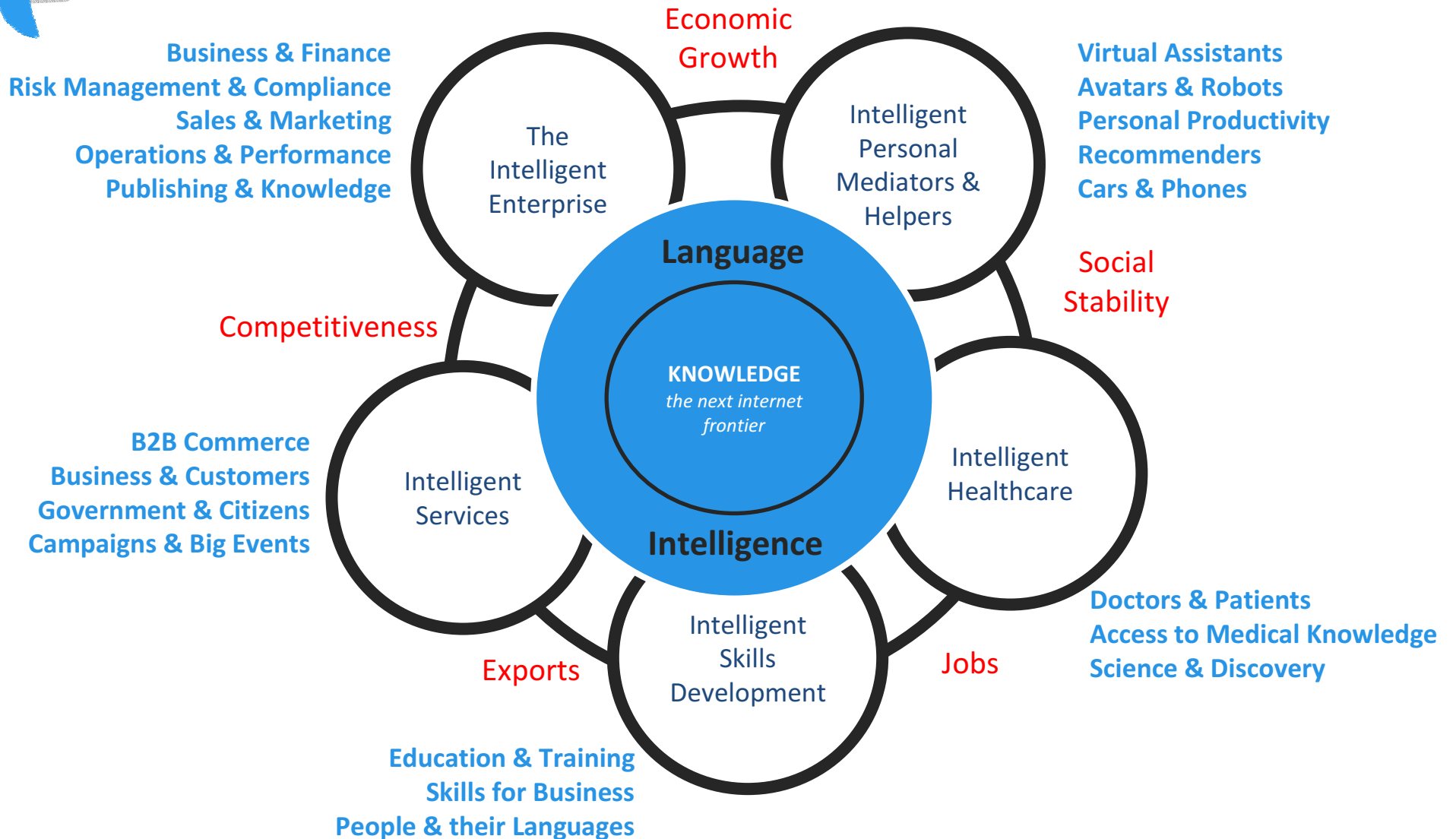
Language = Intelligence

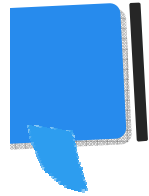




# Where Language Technology adds value

Knowledge: the last Internet frontier

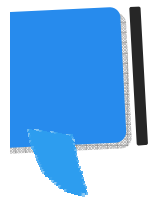




# LT Market Size

The fastest growing ICT Market

- The global LT market is **estimated** to be worth 8 B Euro, and to triple in size over the next five years
- **Speech** has the largest share of the market
- **Translation Technology** represents 2-3% of the language services market, and that share could well double in the next five years
- **Intelligent content** (semantics, search, analytics, contextualisation) will be omnipresent in our IT systems
- LT is probably the **fastest growing segment in IT/software**: it is CURRENTLY growing at 25%/year (outperforming the sw market in general and bucking the downturn because it delivers urgently needed functionality, and exceptional value-for-money/ROI)
- LT represents around **4% of the global software market**

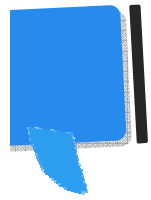


# The problem/opportunity

Language Technology is the key enabling technology par excellence

- In Europe, language is a key last barrier to large-scale digitalisation and valorisation of content/data
- The problem is magnified by the number of languages (63+)
- Language Technology is the solution to the problem
- Providing this solution would entail invaluable spill-over effects for many segments of the economy and society

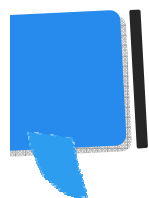




## Knowledge <- Intelligence <- Language Technology

The command of Language is essential to the command of Knowledge

- **Knowledge** is the next Internet frontier
- Command of **Knowledge** will determine social/economic success
- The raw material of **Knowledge** is **Intelligence** (understanding the meaning of things)
- **Intelligence** is encoded in **Language**
- **Language Technology** encompasses the many ways to capture and exploit the **Intelligence** embedded in **Language**
- **Therefore, Knowledge**-based products & services rely, fundamentally, on the ability to process **Language** using **Language Technology**



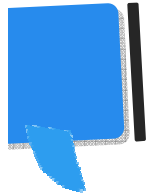
# European Digital Agenda

The missing piece of the puzzle is **Language Technology**



**Language  
Technology**





# European Digital Agenda

A central piece of the puzzle is **Language Technology**

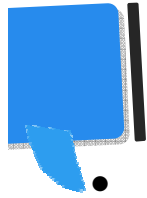
- **LT** is a **key enabler** for the future of the European digital economy and society.
- **LT** includes ***Translation Technology***, a driver for global digital business and export, as well as European digital markets.
- But **LT** is *not just translation!* It is **the** core technology for two additional defining features of the Future Internet.
- **Intelligent Content Technology**: content without “intelligence” (ways to access embedded knowledge) is dead and useless and will have **NO VALUE** in the Future Internet.
- Natural Human Interactivity: the drive for intelligent interfaces is inexorable and ***Speech Technology*** will be the principal new mode of digital interaction, along with touch and gesture.



# Language Technology is **NOT** OPTIONAL

Will Europe participate in the future LT market?

- Every digital product or service either **already has**, or is likely to have, an **LT** component.
- The European Single Digital Market simply **will not exist** without **LT**.
- European industry competes in the global market on **LT-enabled** systems and **will not succeed** without **LT**.
- eGovernment services **cannot be developed and delivered** without **LT**.
- The future of healthcare **is being built** on **LT**.
- **LT** makes a **significant contribution** to the development of digital skills.
- **LT could become the single most pervasive technological enabler since the invention of the Internet.**
- The question is: will Europe participate in the future LT market?



# If Language Technology is so important...

... Why is it invisible?

- Language is everywhere, everybody uses it, everybody is a “language expert” at some level, so language processing seems intuitively “easy”, even uninteresting.
- LT is used “in the background” and is rarely the main functionality that users see.
- Some LT seems “simple” and therefore unimportant, like spell-check; not so, e.g. similar technology is used for innovations like predictive input on devices using gesture.
- Some LT techniques are highly computational (e.g. statistical algorithms), and are not “perceived” to be LT, but they are LT, often integrated with other LT tools.
- There is confusion about AI and semantics: LT encompasses semantic technology, and the traditional discipline of Artificial Intelligence; language is encoded intelligence, and language processing uses intelligence to derive knowledge.
- The most visible application of LT is automatic translation, and translation is a very under-valued service; people expect it to be easy and free, when it is one of the most computationally challenging applications that exists.
- Machine translation is “black box”, people don’t know what’s going on to make it work; this is equally true of intelligent content and speech applications, where the LT functionality is even less visible.
- **People do not realise that LT is at the heart of the most commonly used function on the Internet: Search.**

# Language Technology is important

Why is it an issue for the Digital Agenda?



# The European **Language Technology Industry** is **AT RISK** from lack of coherence

- European **LT** companies, with very few exceptions, are small, and have little market impact.
- There are hundreds of active **LT** companies in Europe, but few grow and develop; most remain “trapped” in their national, linguistic markets.
- The European industry lacks coherence, visibility, or influence in the development of **LT** markets; it simply has no clout.
- The main segments of the European industry (**speech, intelligent content, translation**) operate independently and fail to integrate, cross-pollinate, or exploit commonalities (e.g. sharing and re-use of **LT** tools and resources across different applications and product areas and languages).
- The need to develop multilingual **LT** capacities is critical for Europe, but there are no mechanisms for coordinating unique linguistic expertise in small companies across the Union; they are occasionally “picked off” by global players, but otherwise languish in their linguistic silo’s.
- There is a serious risk of the loss of European **LT expertise** as **LT** innovators and entrepreneurs (trained in European universities) de-camp to other regions where markets are more easily addressed.



# The European **Language Technology Industry** is **AT RISK** from competitive pressures

- The EU has made significant investments in **LT research**, but much of the output has been acquired and exploited by non-EU (mostly US) companies.
- Almost all the results of EC-funded **Speech Technology** research from the last decade are now owned by **one large US company** (with annual sales of \$1.4B in pure speech technology).
- Asian companies are gaining dominance in some areas of **LT**; Indian software companies are actively targeting the EU market in **intelligent content**, building on their market weight in contact centres.
- China has the potential to rival the US in **LT** through substantial Internet R&D commitments, building on their giant Internet user base.
- Europe is at risk of becoming a third-rate player in a market where it has world-class scientific **LT expertise**.
- European **LT** companies do not have the capacity to scale appropriately to address global markets for LT-enabled products and services.
- Non-European companies have a better track-record for addressing **multilingual markets** through localisation (a key application area for **LT**) than European companies.



# Support the European LT Industry

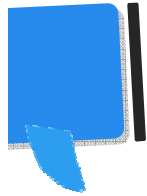
Social & Cultural reasons

## Social reasons

- LT is the greatest enabler of “accessibility” in all its forms, whether linguistic (access to information in any language), pragmatic (access to the right information at the right time), physical (access to digital devices), even political (access to government and public resources); supporting the industry that enables comprehensive accessibility is a social good in itself.
- Without a European LT industry, smaller language communities are guaranteed to be “second class citizens” on measures of accessible information, products and services.
- Indirectly, LT is potentially a great promoter of social cohesion in Europe as it brings equality of access and usability to all citizens, whatever their physical condition or location, or what language they speak.

## Cultural reasons

- European citizens have the right to control the fate of their own languages in the increasingly dominant world of the global Internet; without a native industry that works with their languages, they will lose this control.
- Enabling Europe’s rich cultural resources with “intelligent accessibility” using LT is a gating factor: without European LT, the cultural heritage of Europe either will not be fully digitally available, or will become a resource to be exploited by actors outside the Europe.



# Support the European LT Industry

## Economic reasons

- LT as an industry is estimated to be worth around 8 B EUR globally, and has among the highest growth rates in the ICT sector, reflecting the increasing importance of LT to next-generation digital technologies.
- The ICT market is being transformed (trends including mobile, social/2.0, cloud, crowd, unified publishing and access) in ways that demand the use of LT; the market exists and will be filled – whether European LT companies will have a significant stake in that market is still an open question.
- Europe has the opportunity to compete for a substantial share of the global LT market, driving much-needed exports; intrinsic barriers to accessing those markets must be addressed if this is to happen.
- Without LT, the idea of a Single Digital Market for Europe is a pipedream; cross-language accessibility is a *sine qua non* and can only be delivered with next-generation translation technologies.
- Whether Europeans, or actors from other regions, deliver innovative LT solutions will determine the shape and scope of participation in the digital economy for some regions of Europe, of great significance for their ability to grow economically.

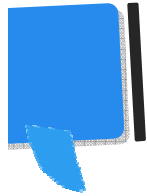




# Support the European LT Industry

## More Economic reasons

- The economic value of LT goes far beyond the LT industry itself, deep into the competitiveness of the European IT/software industry in general, and the strategic competitive advantage to companies that use LT-enabled systems.
- Supporting a native LT industry is a win-win for Europe: innovation is encouraged across the value-chain,
  - from the research groups and startups that develop new and better LT tools and functions
  - to the European software and service companies that develop and embed LT in their products and services
  - to the downstream European industries that use LT-enabled systems for competitive advantage
  - to European customers and end users who get better and more functional products and services , accessible in their own languages using their preferred interfaces.
- If the LT market is abandoned to major players from outside Europe, the exploitation of LT will be increasingly concentrated in global multinationals, squeezing European regional players, even those with good growth potential, to the sidelines; this is exactly what happened in the desktop software market a generation ago.



# Save the European LT Industry

What can be done ?

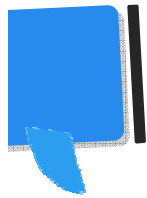
- The European LT Industry is at a critical juncture; the fruits of RTD spending are ripe for exploitation; future R&I must be funded within realistic innovation scenarios that directly engage commercial players.
- To compete *even within the EU*, the European industry needs a proxy for the pure market weight and resources available to their competitors; the need is even greater for global competitiveness; this demands action on a European scale to redress market imbalance.
- The industry needs an LT Innovation Infrastructure, with an **industry** (rather than academic) focus. This could include “Virtual Clusters” to coordinate linguistic islands, and co-sponsorship from the downstream industries that need LT to thrive, or even to survive (software, telecommunications, publishing, media, electronics, pharma, tourism...)
- There is a critical missing ingredient that would enable the European industry to scale, namely shared LT resources, the basic tools for language processing; mega-players can afford to develop and maintain the resources that are used in LT, small European companies cannot.
- A European Language Layer in the Cloud could provide accessible shared LT resources as the basis for scalable, cross-language LT products and services, developed in all languages of the EU, and of global addressable markets.



# European Language Cloud

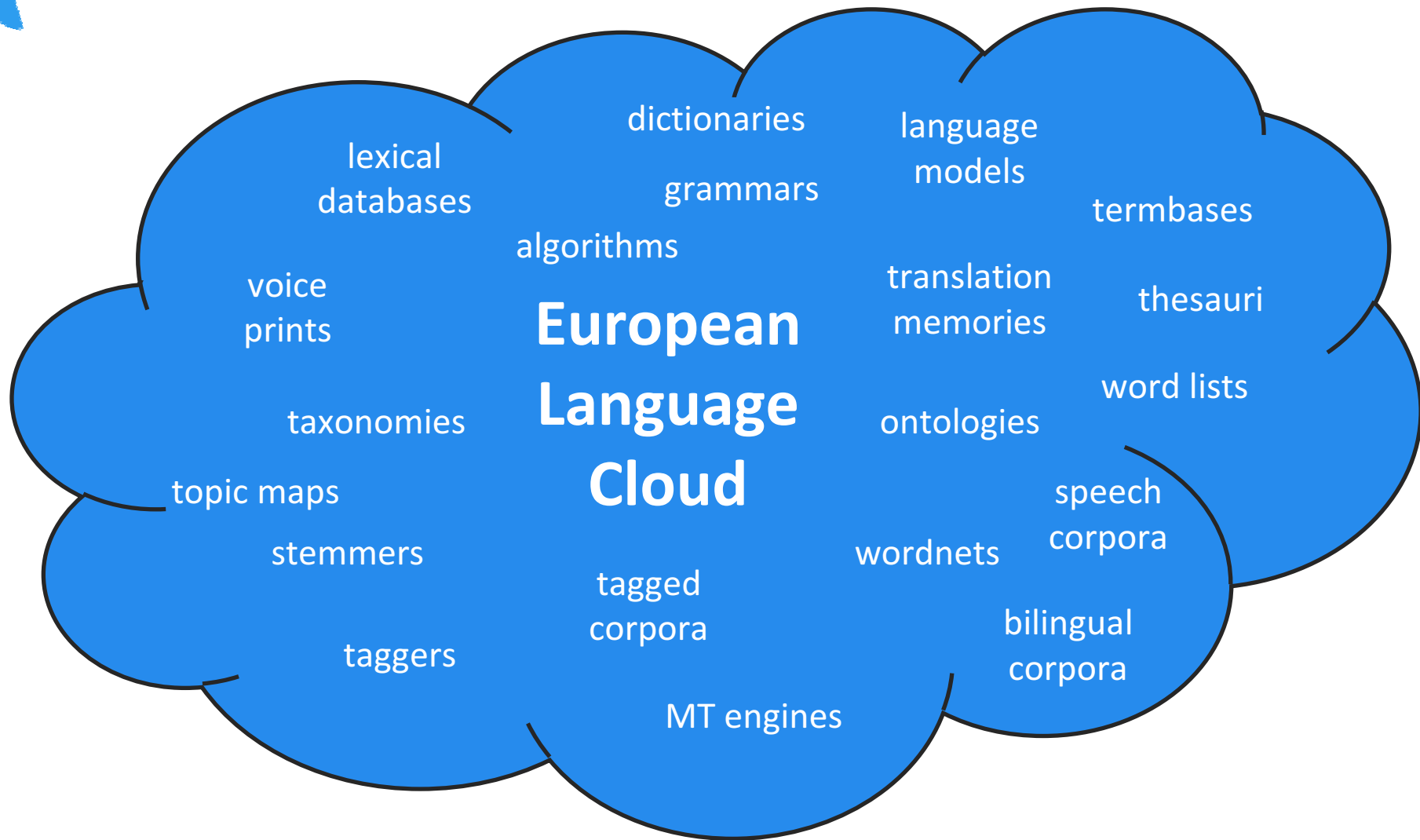
an infrastructure for Language Intelligence providing shared LT resources

European Language Cloud  
the infrastructure of Language  
Intelligence



# European Language Cloud

an infrastructure for Language Intelligence providing shared LT resources





# The way forward ?

Capitalise on a **KEY ASSET!**

- Re-assess the role of LT as a key enabling technology in all policy documents
- Invest an estimated 1B EUR to create “language highways” – European Language Cloud (roughly the cost of a few hundred km of motorways)
- Make a clear distinction between research and innovation. Define a clear set of objectives for each
- Put market-near innovators (SMEs) in the driving seat if the ultimate policy objectives are to encourage new product development, job creation and market expansion
- Use procurement rather than projects for infrastructure (“Cloud”)
- Large scale projects (PEPPOL, STORK, etc.) should include LT layer
- Investment into broadband networks (Connect Europe Facility) should include generic services/middleware and content – and require strong partnerships between telecom operators and innovators
- Support the “europeanisation” of mature national SMEs: facilitate innovative forms of trans-European cooperation between SMEs to overcome fragmentation (partnerships)
- Support SME organisations as key catalysts at European level